图书情报工作
第 65 卷 第 9 期　2021 年 5 月

负责任计量与科学评价趋向*

■ 宋丽萍　王修涵　苑珊珊
天津师范大学管理学院　天津 300387

摘　要：[目的/意义]以负责任计量三大基础框架《旧金山宣言》《莱顿宣言》《计量潮报告》为出发点，就负责任计量视角下科学评价发展趋向进行解读。[方法/过程]基于 Web of Science，借助 VOSviewer 将 2012－2020 年论文和核心作者进行可视化，以反映该领域历时研究热点变化。[结果/结论]基于时间线的关键词、作者叠加视图表明：计量变革正在进行中，并呈现以研究质量为圆心、科研人员为根本、学科规范化为导向、影响力为半径、引用探索为根基、补充计量为参照的发展趋向。结论指出：以负责任计量进行科学评价治理是建立健康学术生态系统的必由之路，而负责任计量的实施意味着对于根深蒂固的激励机制与计量文化的变革。

关键词：负责任计量　科学评价　《旧金山宣言》　《莱顿宣言》　《计量潮报告》
分类号：G250.252
DOI：10.13266/j.issn.0252－3116.2021.09.002

科研评价的世界正在发生变革。2015 年，英国谢菲尔德大学 J. Wilsdon 教授等在《计量潮：研究评价和管理指标作用的独立审查报告》（简称"《计量潮报告》"）中提出负责任计量（responsible metrics）的概念[1]。该报告与《关于科研评价的旧金山宣言》（简称"《旧金山宣言》"）、《莱顿宣言》共同构筑了负责任计量的主体框架。2018 年，J. Wilsdon 又在英国《卫报》上撰文指出：计量潮正在向负责任计量涌动[2]。本文将从上述三大纲领性文献出发，就负责任计量视角下科学评价发展趋向进行分析。

1　相关背景回顾

《计量潮报告》为 2014 年英国基于计量的高等教育评价体系 REF（卓越科研评价框架）首次实施的产物。其中，依据欧盟"负责任研究与创新"的科研治理框架提出了研究评价中合理使用量化指标的思路，即负责任计量，并将其界定为精确数据为基础的稳健性、体现学科间差异的多样性、评价方法的多维性、数据收集和分析过程的透明性、不断更新的自反性，由此诠释了其内涵，并奠定了负责任计量思想的格局[1]。

《旧金山宣言》形成于 2012 年 12 月 16 日在旧金山召开的美国细胞生物学学会（American Society for Cell Biology）第 75 届年会。其主张开发不依赖期刊影响因子的稳健、高效的评估方法，反映出科学共同体纠偏以刊评文、改变科研产出质量评价方式的强烈呼声[3]。

《莱顿宣言》源于 2014 年在荷兰莱顿召开的国际科学技术指标（Science and Technology Indicators，简称 STI）大会上美国佐治亚理工学院公共政策教授 D. Hicks 提出的合理利用科学评价指标的七原则。2015 年 4 月，D. Hicks 与莱顿大学科学技术研究中心执行主任 P. Wouters、副主任 L. Waltman 将其扩充为 10 条，并在《自然》杂志上发表，其中凝练了基于指标的科研评价规范，并预示着科研评价的变革正在酝酿之中[4]。

上述 3 篇纲领性文献代表了科学评价的新思维，构筑了负责任计量的思想空间，绳墨了科学评价的新规范。在其指导下，许多国家科学评价实践中都遵循了负责任计量原则，P. Wouters、L. Waltman、L. Bornmann、M. Thelwall 等计量学家在补充计量（Altmetrics）应用、研究影响力等问题的探索中也将负责任计量原则融入其中[5]。为此，2018 年召开的第 23 届 STI 大会上，L. Waltman 做了题为"负责任计量"的主旨发言。同年，英国负责任计量论坛发布了题为"计量文化转

* 本文系国家社会科学基金一般项目"负责任计量视角下科学评价方法与指标优化研究"（项目编号：18BTQ074）研究成果之一。
作者简介：宋丽萍（ORCID：0000－0003－3056－3720），教授，博士，硕士生导师；王修涵（ORCID：0000－0003－1974－5080），硕士研究生，通讯作者，E-mail：wzxh118@163.com；苑珊珊（ORCID：0000－0002－2736－3333），硕士研究生。
收稿日期：2020－12－24　修回日期：2021－02－09　本文起止页码：16－22　本文责任编辑：徐健

宋丽萍,王修涵,苑珊珊.负责任计量与科学评价趋向[J].图书情报工作,2021,65(9):16-22.

折”的调查报告[6]。可见,负责任计量给科学评价理论与实践带来的改变。

2 2012-2020 历时研究热点可视化

为了呈现近年来负责任计量研究热点及其发展趋势,本文利用 WoS(Web of Science),借助荷兰莱顿大学 N. J. V. Eck、L. Waltman 基于 JAVA 开发的 VOSviewer 将 2012-2020 年论文和核心作者可视化[7]。该软件核心思想是通过共现聚类及其可视化分析,展现知识领域的结构、进化、合作等关系,从而揭示研究热点、作者集团。检索日期为 2020 年 6 月 26 日,检索式为:TS = "responsible metric * " OR TS = "research evaluation * " OR TS = "research assessment * " OR TS = "science policy * " OR TS = " impact factor * " OR TS = "journal impact factor * " OR TS = "citation impact * " OR TS = "altmetric * " OR TS = "article level metric * " OR TS = "scholarly impact * " OR TS = "academic impact * " OR TS = "citescore * " AND Publication years = 2012-2020 and web of science categories = information science library science,检索获得论文 1 778 篇,进而选取基于时间线的关键词、作者叠加视图进行可视化,以反映该领域历时研究热点的变化和发展趋势。

由图 1、图 2 可清晰看出,这一波演化路径为:以期刊影响因子、网络计量探索为起点,逐渐向科研评价、科技政策、研究质量、研究人员研究过渡,并于近期向引用影响力(含学科规范化引文影响力)、补充计量(应用)、合作与创新研究延伸。基于时间线的作者叠加视图则表明 L. Waltman、P. Ingwersen、R. Rousseau、L. Leydesdorff、L. Bornmann、M. Thelwall、H. F. Moed、W. Glanzel 等成为其中核心人物,而 V. Lariviere、K. Kousha、R. Haunschild 成为近期活跃分子。那么,我们应该如何解读上述结果?

3 科学评价发展趋势解读

图 1 勾画出负责任计量背景下科学评价的发展轨迹,并呈现以研究(含研究质量、研究人员、学科差异性)为圆心,以影响力(含引用影响力、社会影响力、补充计量应用)为半径的特质。作为科学计量史上具有

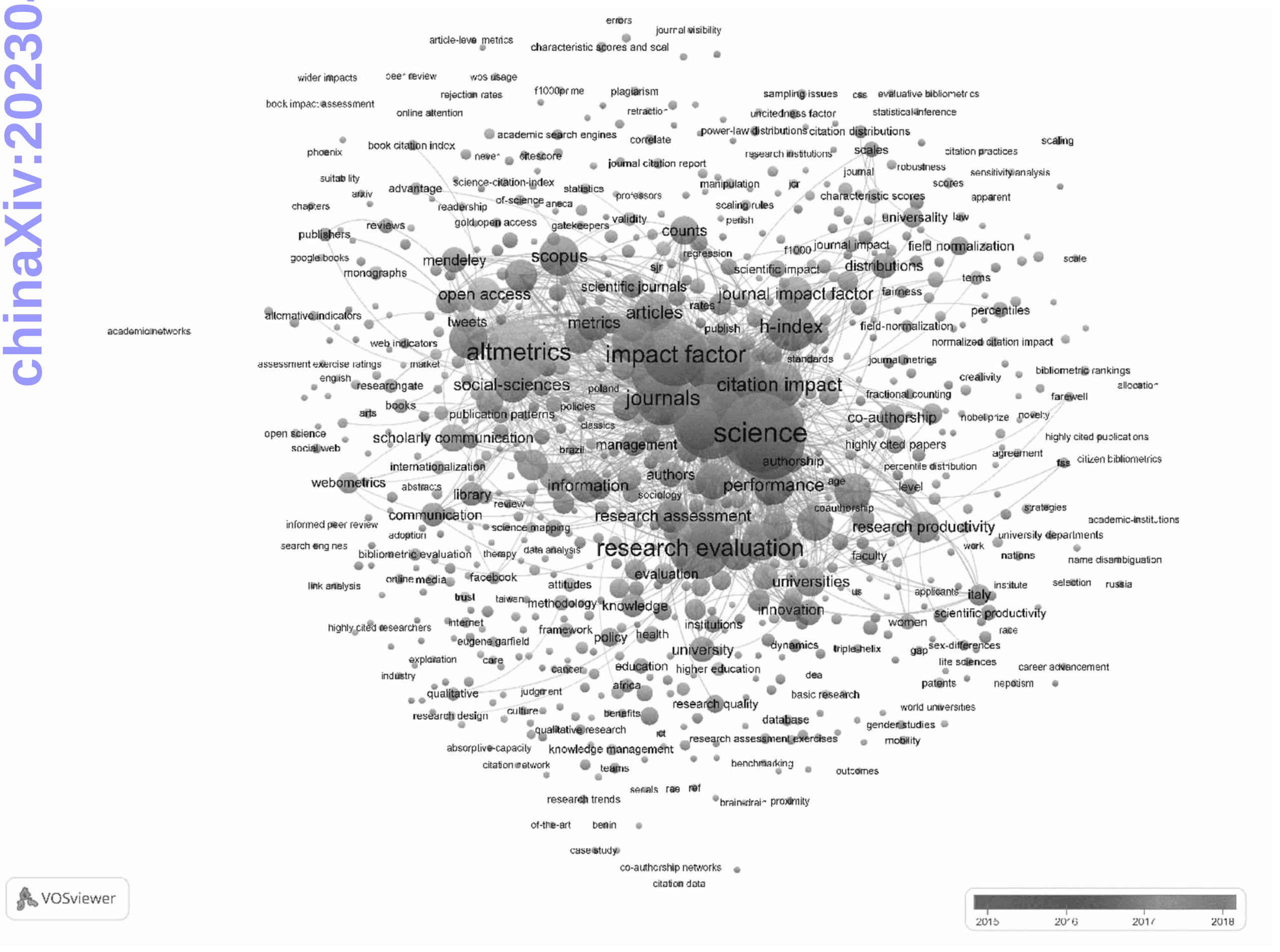


图 1 2012-2020 年负责任计量研究论文关键词叠加聚类视图

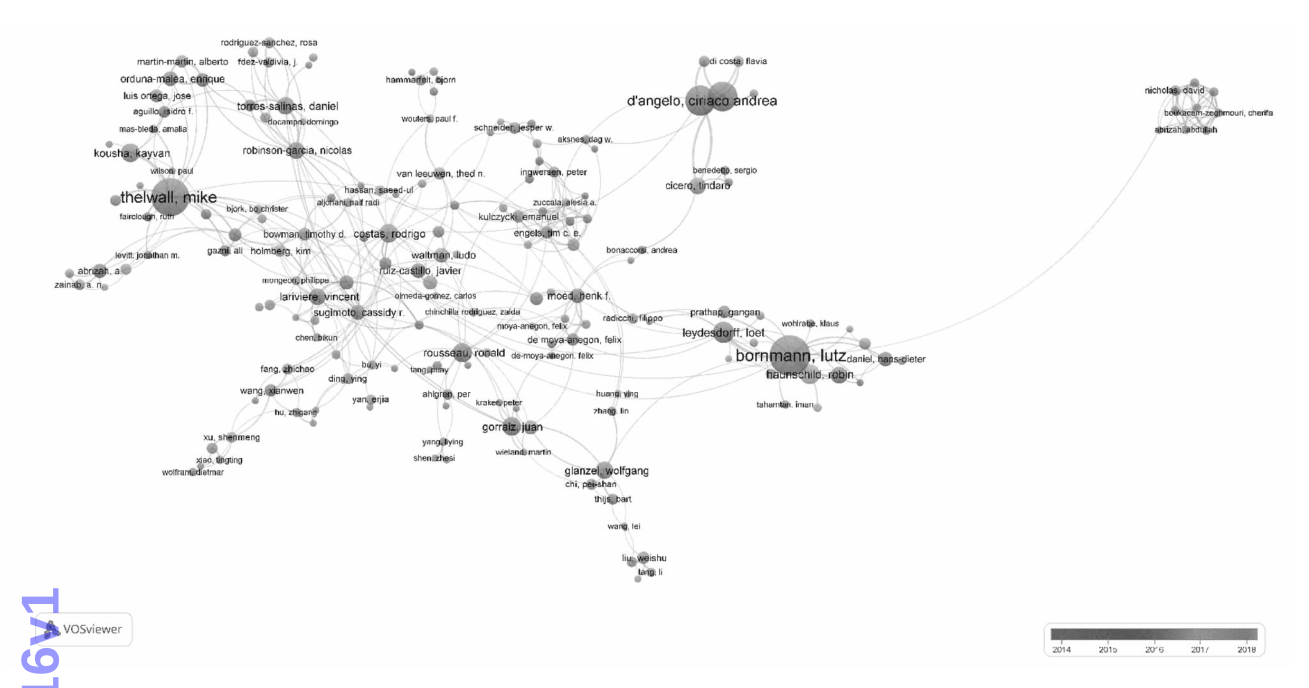


图 2　2012 – 2020 年负责任计量研究论文作者叠加聚类视图

重要意义的事件，其发展轨迹不仅标志着变革的现在进行时，并预示着科学评价的发展趋势，下文将结合近年负责任计量实践对此作出具体分析。

3.1　以研究质量为中心

图 1 中聚类形成的“研究质量”簇及近期科研评价实践说明了负责任计量以研究为圆心的趋势。从根本上来说，作为欧盟“地平线 2020”计划以负责任研究与创新理念进行研究治理在科学评价领域的延伸，负责任计量自出现之日起即与科学研究结下不解之缘。《旧金山宣言》亦旨在建立“以研究为中心”的评估方法；《莱顿宣言》则指出：“评价指标改变研究人员的动机进而改变整个科研系统，对此我们要有充分的预期”。它向人们敲响了警钟，提醒人们审慎对待评价与研究之关系[4]。然而，自 20 世纪后半叶出版商竞争的加剧，异化了影响因子选刊的初衷，致使期刊在原有注册、认证、传播和保存基础上增加第五项功能，即评价，并逐渐演绎为研究论文质量的标签。“以刊评文”“量化至上”、简单粗暴的“计件制”评价方式偏离了科学评价的目标，忽视了评价与知识生产相互作用的复杂方式。由此，学术生态系统要素之研究人员、科学研究、期刊如俄罗斯套娃般被嵌套在基于期刊影响因子的评价体系中，论文成为学术界的“硬通货”，影响因子等同于“论文质量”成为 “无需证明的公理”[8]。捷克和意大利的一些资助机构要求申请人在其出版物旁列出影响因子；比利时用于向大学分配资金的公式包括依据影响因子的加权；巴西教育部系统 Qualis 使用影响因子确定资金分配[9]。对于研究质量的忽视，致使科研人员以牺牲创新性研究为代价不得不摘取“低枝果实”，而不敢提出大胆假设，以求在影响力高的期刊上稳妥地发表。由此 CNS(《细胞》《自然》《科学》)“拜杂志教”盛行，在权威期刊上发表成为个人主要出版策略，科学研究严重脱离了默顿科学精神的轨道，最终科学研究事业为其埋单。究其原因，漠视研究质量的错误导向使然。由此再一次验证了 J. Lane 那句名言：偏颇的计量将产生偏颇的科学[10]。人们不禁质疑：以发表媒介替代研究质量的评价意义何在？

因此，《旧金山宣言》指导委员会主席、英国帝国理工学院结构生物学家 S. Curry 在接受《自然》采访时曾表示，“对于影响因子的使用，都应该和香烟一样附带明确的警示语，其是诸多现行学术评价体系之恶的根源。”[11] 可见，负责任计量力主回归科研质量的真正宿主：科研本身。不仅如此，在 2020 年 3 月召开的英国图书情报科学文献计量分会年会上，REF2021 指导小组主席、英格兰研究部主管 S. Hill 教授在题为“未来研究评价的四个驱动因素”的发言中，将研究与研究文化列入其中两个要素，由此释放了重要信号：以研究与研究质量为圆心的趋势将持续下去[12]。

3.2　以研究人员为根本

在以研究为圆心的场域下，图 1 中在 2017 年初出现“研究人员”聚类簇，由此表明负责任计量对于研究人员（含女性研究人员）的关注。L. Bornmann 曾以“科学界的拟态”深刻揭示了研究人员对科学评价指

宋丽萍，王修涵，苑珊珊. 负责任计量与科学评价趋向[J]. 图书情报工作,2021,65(9):16－22.

标的模拟行为。《计量潮报告》则援引帝国理工学院教授之殇指出:“错误的评价标准扭曲了我们的行为、决定了我们的职业生涯,甚至将导致坎特伯雷教堂大主教所谓大学里新的野蛮。”[1]客观而言,在基于影响因子的科学评价体系中,出版商、资助方、研究机构、研究人员同属“困境囚徒”,但是出版商通过与影响因子捆绑、资助方则通过设计游戏规则占据主导,而研究人员只能屈从于既定评价指标的困囿。简单粗暴科研管理方式下的“指标依赖”意味着出版优于科学发现,于是科研人员抑制作为科学研究原动力之好奇心,为迎合指标而将研究成果指标最大化。统计表明,德国研究人员已经改变自己的出版策略,将论文发表在高影响力英语期刊上以增加论文的被引;澳大利亚在 SCI 论文数量节节攀升的同时,被引出现明显下滑;美国布法罗大学 A. B. Wagner 甚至提出“被引或者出局”(Get Cited or Perish)作为出版或者出局的升级版,并成为科研人员的座右铭。生存的压力、对于影响因子的“痴迷”、不正当激励的响应与博弈助长了科学界的投机取巧,导致了拆分发表、名誉作者、引用卡特尔(引用联盟)和科研造假等学术不端。经济学中古德哈特定律(Goodhart's law)的科学评价版本可谓对这一现象最好的注解:当一个指标变成目标,那么它将势必失效,因为人们开始把它玩弄于股掌之中[13]。

鉴于科学评价、科学共同体、科技政策、科学研究的耦合效应,作为传统科学评价的逆向思维,负责任计量范式下的科学评价加强了与研究人员的对话,将科研人员作为参考系纳入科研评价体系之中。2017 年,科睿唯安、荷兰莱顿大学科技评价中心召集研究人员召开题为“反思影响因子:期刊评价的新途径”的研讨会;2019 年,兰德欧洲基于 3 600 名研究人员的调查形成题为《研究格局的变化及对未来国家研究评估的思考》的报告,以供 REF2021 参考;霍华德休斯医学研究所将北美及欧洲 30 多个研究机构的研究人员聚集一堂,共同探索《旧金山宣言》的实施。如今,作为科学共同体对于科学评价“软约束”的三大纲领性文献,逐渐变成欧盟及澳大利亚、英国等国的硬政策。此番操作意味着对于计量文化、激励机制的探索与思考,并显露出以人为本的负责任计量文化的端倪。

3.3　以学科规范化为导向

伴随负责任计量实践的深入以及以研究为圆心的趋势进一步扩散,图 1 出现学科规范化聚类簇,WoS 文献检索列表中 field-normalization(即学科归一化)亦作为高频词出现,由此折射出负责任计量理论与实践探索中对于跨学科评价的关注。从根本上说,负责任计量中的学术影响力规范化“热”源于 2014 年英国 REF 实施中不同学科比较之需。由于偏态性是引用的特质,即因引用密度(citation density)的差异,被引频次依学科、出版年、文献类型呈高度偏态分布,1979 年 E. Garfield 即指出:不同学科的原始被引频次不能直接进行比较,因此学科规范化成为一种必然[14]。作为规范化方法之一,学科归一化目的即在于,通过将不同变化范围的数值映射到同一量纲下,即[0,1]或者[－1,1]区间内,以消除学科因素给科学计量带来的噪音。图 2 作者叠加视图中以 R. Haunschild 为核心的近期活跃研究作者集团的出现进一步表明负责任计量实践中对于学科规范化算法的探索。2016 年,美国国立卫生研究院 R. Haunschild 提出 RCR(相对引用比),其中基于共引网络动态定义论文所属学科,并实现引用指标的归一化,由此规避了因学科的层级结构及动态发展对于 CNCI(学科规范化引文影响力)等指标的束缚,从而为跨学科,特别是交叉学科研究成果引用影响力评价与比较提供了新的解题思路。毋容置疑,交叉学科研究在促进团队合作、解决挑战性问题、整合零散的学术世界、重塑学术格局方面具有显著的优势,然而目前的悖论是:政府鼓励交叉学科研究,但由于其评价方法的非常规性,交叉学科研究的价值常常受到延迟认可,甚至被漠然置之[15]。2016 年,英国国家学术院院长、世界银行首席经济师 L. Stern 在为 REF2014 所做《斯特恩评论》中强调“交叉学科研究在解决全球社会,经济,生态和政治等复杂问题方面的重要作用”,同时也表示出一丝担忧,认为交叉学科研究在评审中受到不公正的待遇[16]。现今虽然由从事物理学和生物学交叉科学研究的剑桥大学教授 D. A. Donald 组建 REF2021 交叉学科研究评价小组,并承诺在评价中将保持不偏不倚的政策[17],但是 REF2021 提交的相关案例申请仍然远远低于预期,这一切似乎都在拷问科学评价制度。同时,图 1 中 2017 年合作、创新聚簇集的形成、对于 CNCI 等基于先验分类的学科规范化引文影响力的探索、以共引聚类识别研究领域的 RCR 等指标的开发,进一步反映出如何理解和支持多样化和动态的学术研究生态系统、特别是交叉学科研究,成为负责任计量的着力点之一。

3.4　以影响力为半径

图 1 表明负责任计量浪潮下的科学评价在以研究为核心向外蔓延的同时,影响力成为这一波的关键词。检索结果中,1 778 篇文献中 415 篇标题中含有影响

力,L. Waltman的《引用影响指标文献综述》一文在同一出版年论文中以284次高频被引,皆证实这一波对于研究影响力的关注。事实上,以经济学中成本效益理论为准绳,无论是REF2014抑或REF2021,影响力都是其中主旋律。2014年首次以影响力为重心展开评估,并将其界定为:对学术界以外的经济、社会、文化、公共政策、服务、健康、环境或生活质量等方面产生的影响和收益,英国科技部大臣和财政大臣特委托L. Stern为REF2014做了《斯特恩评论》,REF2021实施细则中则进一步加大了影响力的权重[16]。兰德欧洲公司所谓"通过强化影响力使科学、技术和创新给社会和经济带来的收益最大化"的理念一语道破了此种操作的渊源[18]。具体地说,基于负责任计量的科学评价已经被视做提升英国研究能力和质量的关键手段和促进科技进步的杠杆,而实现这一目标的路径即是将影响力划为重点,其中将赌注押在基于影响力的资源分配和机构与个人行为的激励上。

学术影响力与社会影响力代表着影响力评价的两个维度。其中,基于R. K. Merton为代表的规范理论,即"引文即是用来偿还研究债务的货币、代表着对科学研究工作或认知的影响"的论断,学术影响力被量化为引用的"因变量"[14]。与此同时,伴随着网络的发展,影响力向社会影响力扩散并被称之为评价领域的"库恩革命"[19]。为此,负责任计量努力寻求"引用等价物"用于社会影响力的测度。伦敦国王学院与数字科学公司(Digital Science)对于REF2014中7 000个影响案例进行数据挖掘,提取了社会影响力常见量化指标。2018年,兰德欧洲受英格兰高等教育基金管理委员会(HEFCE)之托,发布REF案例分析影响力定量指标准,并开发了由300个问题构成的分析建议包Impact-Finder,以收集研究对于社会、文化和经济的影响。L. Bornmann、R. Haunschild运用聚合效度对案例分析和补充计量在社会影响力测量结果方面的相似度进行比较[20]。研究表明,目前除专利技术影响、医学临床影响两种引用当量的社会影响力测度方法,案例研究是唯一可行的途径。鉴于引用是学术影响力可量度指标已成为学术界的共识,社会影响力的量度作为负责任计量这一波中新的量化元素成为探索中的问题。

3.5 引用探索为准绳

关于引用影响力,图1、图2、英国REF实践、《计量潮报告》以及S. Hill[12]的论断皆表明:无论过去、现在乃至将来,引用都是科学评价的立足点。具体而言,在负责任计量浪潮冲击下,引用依然是科学评价的标准砝码,引用与学术影响力的等价关系依然是科学评价的大前提。究其根源,首先,科学是礼品经济,其价值由其对知识总体的贡献或者对他人的影响力来表征,同时由于作者借助引用表示源自其他研究人员及其成果的影响,因此学术影响力与引用具有了相关性。其次,引用的本质即是知识的流动,即汤森路透首席科学家H. H. Small所谓"引用是信息被利用后留下的路标"[14],它展现了知识单元的离散和重组、继承和创新、演进与升华的复杂过程,由此以引用为科学交流的痕迹、影响力为标识,可以追溯科学发展脉络、优化科学交流、推进科学研究。再者,引文分析可用于识别和绘制研究前沿、定义学科和新兴专业、确定研究计划和项目的跨学科特征和影响。因此,基于科学论文网络揭示科学发展结构,呈现多学科、跨学科关系,从而扶植、鉴别卓越研究,有效指导科学资源配置。从这一意义上说,引用和引文分析与科学评价的总体目标具有较高的契合度。因此,即使S. Woolgar基于"引用是除了科学影响力的许多变量的函数"的论断对引文分析提出质疑[14],M. H. Macroberts指责学术界由于因循保守的"卡桑德拉情结"不肯放弃引文分析,却始终没有动摇引用在科学评价中的地位[21]。

应该指出的是,负责任计量下的科学评价是在引文分析4.0的背景下进行的。除了RCR研究、CNCI等论文层面影响力等问题的探索,全文本文献计量分析成为重头戏,其中通过施引文献与被引文献之间交集的辨识,更深刻地理解科学论文中的引用和动机、更准确地还原科学交流过程,最终更好地服务于科学评价成为今后一段时间内的发展趋向。

3.6 补充计量为参照

图1中补充计量、社媒计量以及爱斯维尔ICSR(国际科学评价研究中心)的多元化科学评价规划证实补充计量在科学评价中拥有一席之地。自2010年J. Priem基于在线科学交流提出补充计量,AM值作为传统文献计量的补充进入关注的视野,同时引发了人们对于计量方式的思考,并催化了论文层面影响力的多维评价[19]。2013年《自然》影响力探索专辑中,科学共同体就影响力的多维性达成共识。《斯特恩评论》亦呼吁对研究影响进行更广泛的定义;与此同时,美国国家信息标准组织(NISO)制定了补充计量使用标准;L. Bornmann、R. Haunschild指出"《莱顿宣言》十原则对于补充计量具有重要指导意义"[22]。J. D. Jenkins则将负责任计量定义为合理使用的引用指标、补充计量指标及其他定量评估手段的综合体,从而为补充计量作为

引用的参照系在负责任评价中的应用定了调子[23]。

然而，尽管一些专家们认为补充计量给学术研究及其影响力计量方式带来了革命性改变，但是 M. Thelwall[24]、H. F. Moed[25]等指出仅 Mendeley 读者统计与引用统计、F1000 同行评议结果具有较高的相关性，可以用于影响力的早期预测；而推特等补充计量指标主要反映的"关注度"，并非影响力。因此，补充计量的有效性是悬而未决的问题。对此，一方面，计量学家通过相关分析检验其与已有指标的契合度，以期解读其含义；另一方面，人们将相对成熟的同被引分析、CNCI 等文献计量方法移植到补充计量中，以期提高其水准，并建立面向科学创造、传播与利用整个工作流的多元化科学评价方法。固然，如同负向引用和过度自引对于引文分析的干扰，数据质量与清洁度将严重影响分析结果，因此如何进行补充计量数据源的筛选与数据的甄别、清洗，并在数据情境分析与统计相关性分析基础上，与同行评议、文献计量有机结合，诠释其内涵，以提高其信度和效度是未来补充计量在负责任计量中应用的关键问题。

4 结语

现今，世界各地的学者以及欧盟委员会专家小组已将科研评价确定为构建全球研究架构的基石，基于影响因子的评价体系已被称为"一种不可再忽视的癌症"甚至是"吞噬科学的数字"[11]，而中国科技政策正在稳步远离基于影响因子的奖励措施。由此可见，以负责任计量进行学术评价治理是建立健康学术生态系统的必由之路，而负责任计量的实施则意味着对于根深蒂固的激励机制和计量文化的变革。

从作为负责任计量发源地——欧洲的科学评价实践以及 VOSviewer 反映的研究热点来看，以《旧金山宣言》《莱顿宣言》《计量潮报告》三大主体框架为核心，呼吁、催化了 21 世纪初科学评价的变革。目前负责任计量文化正在形成之中，并呈现以研究质量为圆心、科研人员为根本、学科规范化为导向、影响力为半径、引用探索为根基、补充计量为参照的发展趋向。虽然由于科学评价的惯性，利益的驱使，负责任计量实施过程一波三折，但是 S. Curry 教授已公布《旧金山宣言实施路线图》、国际研究管理学会研究评价工作组主席 E. Gadd 开发的 SCOPE 框架已具备对评估目的的高度敏感性；ICSR 提出的评估性问询(evaluative inquiry)已具有较强的可操作性；指标工具包 Metrics Toolkit 已在线使用；学术贡献分类 CRediT 已实行；人工智能将帮助我们更完整地复原评价客体的真实面目，而不会因为认识的片面性而重复"盲人与象"的故事。凡此种种都印证着负责任计量实施的脚步。对此，一方面，我们有理由相信：负责任计量不会重蹈"1977 年 E. Garfield 提出防止滥用引文分析后的四十年中，人们对于种种滥用充耳不闻"的覆辙[26]。另一方面，我们必须清醒地意识到，具有稳定性、多维性、多样性、透明性、更迭性的理想的负责任计量尚在探索之中。正如 2020 年 J. Wilsdon 教授在题为"研究评价的新领域"的讲演中所指出的："现今我们朝着正确的方向迈出了第一步，仍需开发更好的指标，付出更多的努力将原则嵌入其中，以建立更敏感的管理框架。"同时，应该指出的是，与定量指标相比，非结构化的同行评议将最大程度上接近被评价对象的真实状态，因此在这个充满不确定性的世界里，基于负责任计量的捷思法和以计量为支撑的知情同行评议的有机结合将成为科学评价的未来走向[27]。

参考文献：

[1] WILSDON J，ALLEN L，BELFIORE E，et al. The metric tide：report of the independent review of the role of metrics in research assessment and management[M]. London：SAGE Publications Ltd，2015.

[2] WILSDON J. Has the tide turned towards responsible metrics in research? [N]. The guardian，2018－07－10(1).

[3] San Francisco declaration on research assessment [EB/OL]. [2020－09－23]. http://www.ascb.org/dora/.

[4] HICKS D，WOUTERS P，WALTMAN L，et al. Bibliometrics：the Leiden manifesto for research metrics[J]. Nature，2015，520(7548)：429－431.

[5] CASSIDY R S. Theories of informetrics and scholarly communication[M]. Berlin：De Gruyter，2016.

[6] WILSDON J. UK progress towards the use of metrics responsibly：three years on from The Metric Tide report. [EB/OL]. [2021－03－23]. https://dera.ioe.ac.uk/31945/1/UK% 20progress% 20towards% 20the% 20use% 20of% 20metrics% 20responsibly% 2010072018.pdf.

[7] 李杰. 科学知识图谱原理及应用——VOSviewer 和 CitNetExplorer 初学者指南[M]. 北京：高等教育出版社，2018.

[8] GUÉDON J C，KRAMER B，LAAKSO M，et al. Future of scholarly publishing and scholarly communicate on：report of the expert group to the European commission [EB/OL]. [2020－09－06]. https://op.e-ur-opa.e-u/en/ publication-detail/-/publication/464477b3-2559-11e9-8d04-01aa75ed71a1.

[9] STEPHAN P，VEUGELERS R，WANG J. Reviewers are blinkered by bibliometrics[J]. Nature，2017，544(7651)：411－412.

[10] LANE J. Let's make science metrics more scientific[J]. Nature，2010，464(7288)：488－489.

[11] CURRY S. Let's move beyond the rhetoric: it's time to change how we judge research[J]. Nature, 2018, 554(7691): 147.
[12] HILL S. Four trends shaping the future of research evaluation[EB/OL]. [2020-09-06]. https://figshare.c-om/articles/Four_trends_shaping_the_future_of_research_evaluation/11918130/1.
[13] WOUTERS P, SUGIMOTO C R, LARIVIÈRE V, et al. Rethinking impact factors: better ways to judge a journal[J]. Nature, 2019, 569(7758): 621-623.
[14] WALTMAN L. A review of the literature on citation impact indicators[J]. Journal of informetrics, 2016, 10(2): 365-391.
[15] VAN NOORDEN R. Interdisciplinary research by the numbers [J]. Nature, 2015, 525(7569): 306-307.
[16] STERN N. Lord Stern's review of the research excellence framework: call for evidence [EB/OL]. [2020-09-08]. https://core.ac.uk/download/pdf/74380480.pdf.
[17] Research excellence framework 2021 [EB/OL]. [2020-09-08]. https://www.ref.ac.uk/.
[18] PARKS S. The changing research landscape and reflections on national research assessment in the future [EB/OL]. [2020-09-08]. https://re.ukri.org/documents/2019/rand-europe-full-report/.
[19] BORNMANN L. Scientific revolution in scientometrics: The broadening of impact from citation to societal[M]// CASSIDY R S. Theories of informetrics and scholarly communication[M]. Berlin: De Gruyter. 2016: 347-359.
[20] BORNMANN L, HAUNSCHILD R. Do altmetrics assess societal impact in a comparable way to case studies? an empirical test of the convergent validity of altmetrics based on data from the UK research excellence framework (REF)[J]. Journal of informetrics, 2019, 13(1): 325-340.
[21] MACROBERTS M H, MACROBERTS B R. The mismeasure of science: Citation analysis[J]. Journal of the Association for Information Science & Technology, 2017, 69(3): 474-482.
[22] BORNMANN L, HAUNSCHILD R. To what extent does the Leiden manifesto also apply to altmetrics? [J]. A discussion of the manifesto against the background of research into altmetrics, 2016, 40(4): 529-543.
[23] DAVID J. What are responsible metrics [EB/OL]. [2020-09-06]. http://www.open.ac.uk/blogs/the_orb/? p=3242.
[24] THELWALL M. Measuring societal impacts of research with altmetrics? common problems and mistakes[EB/OL]. [2020-09-08]. https://onlinelibrary.wiley.com/doi/pdfdirect/10.1111/joes.12381.
[25] MOED H F. Applied evaluative informetrics[M]. Berlin: Springer International Publishing, 2017.
[26] LARIVIERE V, SUGIMOTO C R. The journal impact factor: a brief history, critique, and discussion of adverse effects[C]// GLÄNZEL W, MOED H F. Springer handbook of science and technology indicators. Berlin: Springer, 2019: 3-24.
[27] BORNMANN L, HUG S, MAREWSKI J N. Bibliometrics-based heuristics: what is their definition and how can they be studied? [EB/OL]. [2020-09-08]. https://arxiv.org/abs/1810.13005v1.

作者贡献说明：
宋丽萍：论文选题与撰写；
王修涵：数据分析；
苑珊珊：数据搜集。

Responsible Metrics and the Trend of Research Evaluation

Song Liping　Wang Xiuhan　Yuan Shanshan
School of Management, Tianjin Normal University, Tianjin 300387

Abstract: [**Purpose/significance**] Based on the three fundamental frameworks of responsible metrics: *San Francisco Declaration on Research Assessment*, *Leiden Manifesto for Research Assessment* and *the Metric Tide*, the paper interprets the development trend of research evaluation from the perspective of responsible metrics. [**Method/process**] The articles and core authors from 2012 to 2020 in Web of Science were used and visualized with VOSviewer, to reflect the changes of diachronic research hotspots in this field. [**Result/conclusion**] The timeline and overlay views for keywords and authors illustrate that the revolution in research metrics is in progress, and present the development trend with research quality as the center of the circle, researchers as foundation, field-normalization as goal orientation, research impact as radius, citation exploration as root, and altmetrics as reference. The conclusion points out that research evaluation ruled by responsible metrics is the only way to establish a healthy academic ecosystem, however, it is clear that the revolution for deep-rooted academic ecosystem and incentive mechanism is needed.

Keywords: responsible metrics　research evaluation　*DORA*　*Leiden Manifesto*　*the Metric Tide*